

## Predictions shape confidence in the right inferior frontal gyrus

Article (Accepted Version)

Sherman, Maxine T, Seth, Anil K and Kanai, Ryota (2016) Predictions shape confidence in the right inferior frontal gyrus. *Journal of Neuroscience*, 36 (40). pp. 10323-10336. ISSN 0270-6474

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/63665/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# Predictions Shape Confidence in Right Inferior Frontal Gyrus

 **Maxine T. Sherman**<sup>1,2</sup>  **Anil K. Seth**<sup>1,3</sup> and  **Ryota Kanai**<sup>1,2,4,5</sup>

<sup>1</sup>Sackler Centre for Consciousness Science, University of Sussex, Brighton, BN1 9QJ, United Kingdom, <sup>2</sup>School of Psychology, University of Sussex, Falmer, BN1 9QH, United Kingdom, <sup>3</sup>Department of Informatics, University of Sussex, Falmer, BN1 9QJ, United Kingdom, <sup>4</sup>Department of Neuroinformatics, Araya Brain Imaging, Tokyo 102-0093, Japan, and <sup>5</sup>YHouse Inc., New York, New York 10023

It is clear that prior expectations shape perceptual decision-making, yet their contribution to the construction of subjective decision confidence remains largely unexplored. We recorded fMRI data while participants made perceptual decisions and confidence judgments, manipulating perceptual prior expectations while controlling for potential confounds of attention. Results show that subjective confidence increases as expectations increasingly support the decision, and that this relationship is associated with BOLD activity in right inferior frontal gyrus (rIFG). Specifically, rIFG is sensitive to the discrepancy between expectation and decision (mismatch), and higher mismatch responses are associated with lower decision confidence. Connectivity analyses revealed expectancy information to be represented in bilateral orbitofrontal cortex and sensory signals to be represented in intracalcarine sulcus. Together, our results indicate that predictive information is integrated into subjective confidence in rIFG, and reveal an occipital-frontal network that constructs confidence from top-down and bottom-up signals. This interpretation was further supported by exploratory findings that the white matter density of right orbitofrontal cortex negatively predicted its respective contribution to the construction of confidence. Our findings advance our understanding of the neural basis of subjective perceptual processes by revealing an occipitofrontal functional network that integrates prior beliefs into the construction of confidence.

**Key words:** confidence; expectation; fMRI; inferior frontal gyrus; perception; prediction error

## Significance Statement

Perceptual decision-making is typically conceived as an integration of bottom-up and top-down influences. However, perceptual decisions are accompanied by a sense of confidence. Confidence is an important facet of perceptual consciousness yet remains poorly understood. Here we implicate right inferior frontal gyrus in constructing confidence from the discrepancy between perceptual judgment and its prior probability. Furthermore, we place right inferior frontal gyrus within an occipitofrontal network, consisting of orbitofrontal cortex and intracalcarine sulcus, which represents and communicates relevant top-down and bottom-up signals. Together, our data reveal a role of frontal regions in the top-down processes enabling perceptual decisions to become available for conscious report.

## Introduction

Perception is increasingly being seen as an active process, in which current or future sensory states are inferred from predic-

tive information (Engel et al., 2001; Lee, 2002; Bar, 2007; Beck and Kastner, 2009; Fiser et al., 2010; Gilbert and Li, 2013). These predictions can be modeled in Bayesian terms as prior beliefs, which bias perceptual inference toward solutions that are a priori more likely in a given context (Bülthoff et al., 1998; Seriès and Seitz, 2013; Trapp and Bar, 2015). Predictions, or priors, can have striking effects on perception, especially under high sensory uncertainty. For example, ambiguous rotational motion can be subjectively disambiguated by prior exposure to rotation direction, resulting in the perception of a rotation direction despite none existing in the physical stimulus (Maloney et al., 2005). In laboratory conditions, these behavioral effects of prediction are typically accompanied by increases in BOLD amplitude, ERP amplitude and evoked gamma power, both over sensory (Egner

Received April 1, 2016; revised Aug. 20, 2016; accepted Aug. 22, 2016.

Author contributions: M.T.S., A.K.S., and R.K. designed research; M.T.S. performed research; M.T.S. and R.K. analyzed data; M.T.S., A.K.S., and R.K. wrote the paper.

This work was supported by the Dr. Mortimer and Theresa Sackler Foundation (which supports the Sackler Centre for Consciousness Science) to M.T.S. and A.K.S., the Japan Science and Technology Agency to R.K., and the School of Psychology, University of Sussex to M.T.S.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Maxine T. Sherman, School of Psychology, Pevensey Building, University of Sussex, Falmer, BN1 9RH, UK. E-mail: maxinesherman@gmail.com.

DOI:10.1523/JNEUROSCI.1092-16.2016

Copyright © 2016 the authors 0270-6474/16/360001-15\$15.00/0

et al., 2010; Wacongne et al., 2011; Jiang et al., 2013; Bauer et al., 2014; Kouider et al., 2015) and decision-related (Bubic et al., 2009) brain regions. These neural responses are typically referred to as a “prediction error” response profile and taken to represent the discrepancy between internal templates and perceptual content.

The perceptual content that forms the basis of our visual experience is accompanied by a degree of subjective confidence. Confidence reflects the estimated accuracy of a perceptual choice and can be seen as a gate for postperceptual processes, such as learning and belief-updating (Nassar et al., 2010; Yeung and Summerfield, 2012). However, while subjective confidence is an integral part of perceptual experience that is easy to probe in human subjects (Seth et al., 2008; Sandberg et al., 2010; Overgaard and Sandberg, 2012; Fleming and Lau, 2014; Wierchoń et al., 2014), its construction remains poorly understood.

It is clear that confidence increases with evidence in support of the decision (Fetsch et al., 2014b; Gherman and Philastides, 2015; Hebart et al., 2016). Decision and subjective confidence are thought to evolve together until the first-order, objective decision has been made (Ratcliff and Starns, 2009; Kepecs and Mainen, 2012). Accordingly, there exists strong evidence for a common sensory signal underlying both types of report (Kiani and Shadlen, 2009; Fetsch et al., 2014a; Kiani et al., 2014). However, influences on objective decisions are not always reflected in subjective aspects of decision-making. For example, evidence for the unchosen perceptual inference is underweighted in confidence judgments (Zylberberg et al., 2012; Maniscalco et al., 2016).

Although many studies have investigated the role of top-down prior expectations on objective decision-making, surprisingly little research has investigated the role of such priors on confidence. Converging behavioral evidence indicates that subjective confidence increases with prior evidence in favor of the associated choice (Meyniel et al., 2015a; Sherman et al., 2015). This suggests that the construction of confidence may involve a comparison process between decision and prior, yet the neural substrates accompanying this comparison remain unexplored.

Here we aimed to identify brain regions in which prior perceptual expectations are integrated into confidence judgments. Based on aforementioned previous work, we reasoned that confidence should be high when decisions are supported by prior knowledge (i.e., when the discrepancy between expectation and perceptual decision is low). We therefore sought to identify brain regions that: (1) are sensitive to both “prediction error” and confidence, and (2) in which prediction error and confidence are negatively correlated. In such a region, confidence would be associated with the mismatch between internal templates and perceptual contents.

We further hypothesized that regions found to integrate prior expectations into confidence judgments should be functionally connected with two information sources: one representing the decision evidence or sensory information and one that represents the prior expectation. As confidence increasingly depends on prior expectations, functional connectivity (FC) with the source of the priors should increase. Similarly, when confidence is less dependent on priors, FC with sensory regions should increase.

## Materials and Methods

**Participants.** The study was approved by the Brighton and Sussex Medical School Research Governance and Ethics Committee. Twenty-four healthy, English-speaking and right-handed subjects were tested (age 19–34 years, mean 25 years; 13 females). Data from 5 participants were excluded: 2 for whom thresholding failed (Gabor hit rate = 2%, visual

search  $d' = -0.1$ ); one who revealed abnormal vision only after scanning (whose estimated contrast thresholds were accordingly  $>2$  SD from the mean); one for excessive head movement in the scanner, such that their T1 scan was unusable; and one for failing to respond on 33% of trials (relative to a mean of 3%). This left 19 participants with normal or corrected-to-normal vision for analysis. All participants gave informed, written consent and were reimbursed £50 for their time.

**Procedure.** The experiment was conducted over three sessions, at least 2 h apart (no participant completed all three on a single day). In Session 1, informed consent was obtained. Participants were trained on all tasks before scanning. This consisted of on-screen instructions, followed by a minimum of 10 practice trials of each task. Participants were encouraged to continue training until the task was well understood and response mappings learned. Next, three separate staircase procedures were run to equate detection accuracy across levels of attention and across participants. These were completed in the scanner without acquiring EPIs. Finally, two 17 min runs of experimental trials were completed while EPI scans were acquired.

Session 2 did not include a training component but was otherwise identical to Session 1. Session 3 consisted of 10 min for T1 acquisition, 15 min of retinotopy (data from which are not used in this paper), and, time permitting, one more experimental run (8 participants completed this). Two participants were only able to complete three runs in total, and two participants completed a total of six runs. This variation in number of runs was a result of scanner availability.

Once the three sessions had been completed participants were compensated for their time and debriefed.

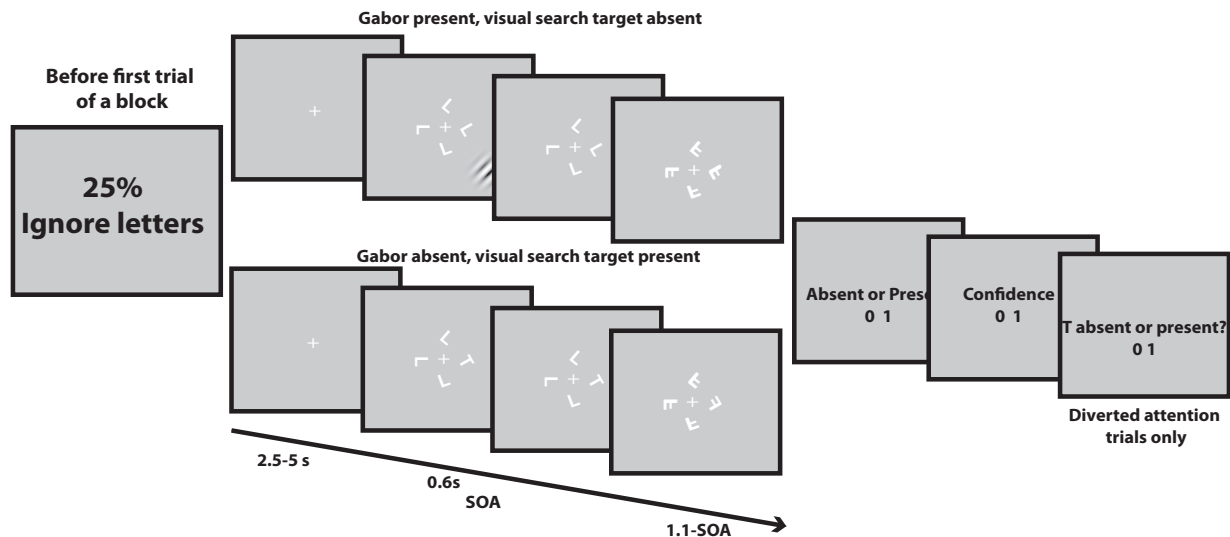
**Experimental design.** The paradigm used in the present study was adapted from a previously reported design (Sherman et al., 2015). The visual display was identical in all sections of the sessions (training, staircase, and experimental). It consisted of a central visual search array and the presence or absence of a to-be-detected Gabor patch in the periphery (Fig. 1; Trial sequence).

In experimental trials, the principal task was Gabor detection. Two factors were orthogonally manipulated: prior probability of Gabor presence and attention. Expectations were manipulated blockwise by manipulating the percentage of trials on which a Gabor was present within the block and informing participants of this probability: a 25% condition induced an expectation of Gabor absence; a 75% condition induced an expectation of Gabor presence; and a 50% condition acted as a control (flat prior). Attention was manipulated by instructing participants to either perform or ignore a visual search task, presented concurrently with the Gabor target. This task consisted of detecting target ‘T’s among an array of distracter ‘L’s. Performing both tasks concurrently diverted attention from the Gabor detection task, allowing us to separate effects of expectation from those of attention.

There were 12 trials in each condition and each condition occurred once per scanning run, in fully counterbalanced order. Participants were informed of both the expectation and attention condition before each experimental block began via the presentation of an instruction screen presented for 10 s (Fig. 1). Participants were instructed to always maintain fixation at a central cross.

**Trial sequence.** The trial sequence was identical for training, staircase, and experimental trials and is shown in Figure 1. These sections differed only in task instructions and response prompts (see Experimental design). Trials began with a white fixation cross of random duration between 2.5 and 5 s. Next, a visual search array appeared, which consisted of seven letters ( $1.6^\circ \times 2.4^\circ$ ): all white, capital ‘L’s (50% chance), or a white, capital ‘T’ replacing an ‘L’ (50% chance). All letters were equidistant from fixation (eccentricity  $4.3^\circ$ ) and took an independently random orientation. These were subsequently masked by a matching array of ‘F’s to increase task difficulty. In total, the visual search array was present for 1.1 s. The stimulus onset asynchrony (SOA) between target and masking arrays was titrated for each participant to achieve 79% accuracy (see Staircases).

On some trials, a near-threshold ( $\sim 4\%$  contrast under full attention and  $\sim 6\%$  under diverted attention) peripheral (eccentricity  $8^\circ$ ) Gabor patch (orientation =  $135^\circ$ , phase  $45^\circ$  on 50% of trials,  $225^\circ$  on 50% of trials,  $sf = 3c/^\circ$ , Gaussian SD =  $0.45^\circ$ ) was also presented. On



**Figure 1.** Trial sequence. Blocks began with instructions signaling the expectation and attention condition. In this example, the participant should expect target absence and ignore the concurrent visual search task (i.e., allocate full attention to Gabor detection). On each trial, a visual search target T was either absent (top) or present (bottom) with 50% probability. On each trial, a target Gabor was either present (top) or absent (bottom). The probability of Gabor presentation changes with experimental condition. Response cues followed the offset of the stimuli. Staircase trials were identical, except that there was no condition-specific instruction at the beginning and confidence reports were not requested.

these trials, the Gabor target appeared simultaneously with visual search array onset. To minimize attentional capture, the Gabor was presented over 0.6 s in a Gaussian envelope over time so that it had a gradual onset and offset. Stimulus contrast was titrated to equate performance across levels of attention and participants at 79% accuracy (see Staircases).

The interval between offset of the masking array and onset of response prompts was jittered during experimental trials. The aim here was to minimize motor cortex activity reflecting response anticipation. Jitter was randomly selected from the discrete values 1.3 s:0.3 s:3.1 s.

Response prompts were presented at the end of the trial. The first prompt asked whether the Gabor had been presented. “Absent”/“Present” responses were recorded by pressing the outer left/right key. This prompt was presented on all trials, except those of the visual search staircase procedure (only visual search performed). The second prompt asked whether participants guessed (inner left) or were confident (inner right) in their Gabor detection response (not presented during staircases). The third prompt was only presented on trials where participants performed the visual search task. This asked whether the visual search target ‘T’ was absent (outer left) or present (outer right). Response prompts remained onscreen for 2 s, and responses were coded as missed trials if no response was given within the allowed time.

**Staircases.** Before each experimental session, three separate adaptive 1-up-3-down psychophysical staircase procedures (nine reversals) were completed in the scanner. The aim of these staircases was (1) to equate Gabor detection and visual search performance across subjects and (2) to equate Gabor detection performance across levels of attention. We did not equate Gabor detection performance across levels of expectation because expectation does not tend to affect detection sensitivity (Kok et al., 2012a; Morales et al., 2015; Sherman et al., 2015).

Staircase trials were identical to those in experimental trials (see Trial structure), except: there was no manipulation of attention or expectation; the Gabor was always present, but randomly oriented either 45° to the left or to the right; the Gabor task was left/right orientation discrimination instead of yes/no target detection; and confidence ratings were not requested.

We used orientation discrimination instead of Gabor detection so that the Gabor would be present on every trial, enabling effective titration. Although this procedure might not lead to precisely 79% correct on the Gabor detection task, performance on this task should still be equated

across levels of attention and across subjects, which was the main goal of the staircasing procedure.

Staircase 1 titrated Gabor contrast to achieve 79% accuracy under full attention. Initial contrast was 1.5%. The visual search array was masked after 0.5 s. Participants were instructed to ignore the visual search array but still fixate centrally.

Staircase 2 titrated the SOA between the visual search array and masking array to set performance at 79% (in the visual search task). Initial SOA was 500 ms. Participants ignored the Gabor orientation task and only performed the visual search task. The (ignored) Gabor was presented at the contrast acquired in Staircase 1.

Staircase 3 titrated Gabor contrast to achieve 79% accuracy in Gabor orientation discrimination under diverted attention (i.e., under dual-task conditions). Initial contrast was set at that obtained in Staircase 1 (under single-task conditions) and visual search SOA was set at the value obtained by Staircase 2. Here, participants performed both the Gabor and the visual search tasks. The visual search SOA was set at the value obtained in the previous staircase and initial contrast was set at that obtained in the first. Gabor contrast was titrated over the course of the staircase to obtain contrast thresholds under diverted attention.

**Statistical analyses.** Gabor detection sensitivity and decision threshold were quantified by computing signal detection theoretic measures  $d'$  and  $c$ , respectively. These are computed by classifying trials as hits ( $h$ ), misses ( $m$ ), false alarms ( $fa$ ), or correct rejections ( $cr$ ). Hit rate ( $HR$ ) and false alarm rate ( $FAR$ ) are defined as follows:

$$HR = \frac{\sum h}{\sum h + \sum m} \text{ and } FAR = \frac{\sum fa}{\sum fa + \sum cr}$$

Where  $Z$  is the inverse cdf of the standard normal distribution, detection sensitivity  $d'$ , and decision threshold  $c$  are defined as follows:

$$d' = Z(HR) - Z(FAR) \text{ and } c = -\frac{Z(HR) + Z(FAR)}{2}$$

Confidence was computed by calculating the proportion of trials on which each subject reported “confident.” We did not use the Type 2 signal detection theory measure of confidence threshold Type 2  $C$  because it is an unprincipled measure (Galvin et al., 2003).

Behavioral and follow-up statistical tests were run on JASP (Love et al., 2015). When the null hypothesis was predicted, Bayesian  $t$  tests and repeated-measures ANOVAs used the JASP default Cauchy prior of



$Y = 0.7$ , centered on zero. All results presented were robust to reasonable adjustments of this value. Bayes factors  $>1/3/10/100$  are, respectively, interpreted as showing insensitive/moderate/strong/very strong evidence for the alternative hypothesis (Kass and Raftery, 1995). Bayes factors less than the reciprocal of these values are given the same labels but refer to the null hypothesis.

Unless otherwise stated, all repeated-measures ANOVA results met the assumption of sphericity. When sphericity was violated, corrected degrees of freedom and  $p$  values were used. The Greenhouse-Geisser correction was used for small violations ( $\epsilon < 0.75$ ) and the Huynh-Feldt correction for large violations ( $\epsilon > 0.75$ ).

**fMRI acquisition and preprocessing.** Functional T2\*-sensitive EPIs were acquired on a Siemens Avanto 1.5T scanner. Axial slices were tilted to minimize signal dropout from frontal and occipital cortices. 34 2 mm slices with 1 mm gaps were acquired (TR = 2863 ms, TE = 50 ms, FOV = 192 mm  $\times$  192 mm, matrix = 64  $\times$  64, flip angle = 90°). Full brain T1-weighted structural scans were acquired on the same scanner and were composed of 176 1-mm-thick sagittal slices (TR = 2730 ms, TE = 3.57 ms, FOV = 224 mm  $\times$  256 mm, matrix = 224  $\times$  256, Flip angle = 7°) using the MPRAGE protocol.

Each functional run lasted 17 min. The first four functional volumes of each run were treated as dummy scans and discarded. Images were processed using SPM8 software (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). Images were preprocessed using standard procedures: anatomical and functional images were reoriented to the anterior commissure; images were slice-time corrected with the middle slice used as the reference; and EPIs were aligned to each other and coregistered to the structural scan by minimizing normalized mutual information. Next, EPIs were spatially normalized to MNI space using parameters obtained from the segmentation of T1 images into gray matter (GM) and white matter (WM). Finally, spatially normalized images were smoothed with a Gaussian smoothing kernel of 8 mm FWHM.

**fMRI statistical analysis.** At the participant level, BOLD responses were time-locked to the onset of the visual search array (which appeared at the same time as the Gabor, if present), enabling us to examine BOLD responses to both target present and target absent trials. BOLD responses were modeled in a GLM with regressors and their corresponding temporal derivatives for each combination of the following factors: Attention (full, diverted), Expectation (25%, 50%, and 75%), Stimulus (Gabor present, Gabor absent), Report (yes, no), and Confidence (confident, guess). If a certain combination of factors had no associated trials for a particular participant, that regressor was removed from the participant's first-level model and contrast weights were rescaled.

The reliability of the regression weights was maximized by entering data from all runs and sessions together, increasing the trial count per regressor. To avoid smearing artifacts, no bandpass filter was applied. Instead, low-frequency drifts were regressed out by entering WM drift (averaged over the brain) as a nuisance regressor (Law et al., 2005). Nuisance regressors representing the experimental run and six head motion parameters were also included in the first-level models.

Comparisons of interest were tested by running one-sample  $t$  tests against zero at the participant level, then running group-level paired  $t$  tests on those one-sample contrast images. Unless otherwise stated, all contrasts at the group level were run with peak thresholds of  $p < 0.001$  (uncorrected) and corrected for multiple comparisons at the cluster level using the FDR method.

We wanted to control for possible confounds between reaction speed and confidence (Petrucci and Baranski, 2003) which correlate (see e.g., Grinband, Hirsch, and Ferrera, 2006), and between individual or condition-wise differences in Gabor contrast and confidence (which correlate) (Rahnev et al., 2011). To do this, a control GLM was computed. Here, each regressor was parametrically modulated by both Gabor contrast (linear modulation) and reaction time (quadratic modulation). By design, this model controlled for (1) reaction time-confidence associations and (2) across-subject and across-condition differences in Gabor contrast. The Results section reports analyses on our main model (the model without parametric modulators) because the control model had a fourfold increase in number of regressors, reducing statistical power. Nonetheless, all GLM analyses were replicated under our control model

when using a peak threshold of  $p < 0.005$ . Crucially, all results under right inferior frontal gyrus (rIFG) were also replicated when using a peak threshold of  $p < 0.001$ .

Functional ROIs were extracted using the MarsBaR 0.42 toolbox (<http://marsbar.sourceforge.net/download.html>). Anatomical areas were labeled using the SPM Anatomy toolbox (Eickhoff et al., 2005). Brodmann areas were identified using MRIcro (Rorden and Brett, 2000). Results of whole-brain analyses were plotted onto glass brains using MATcro (now called MRIcroS, <https://www.nitrc.org/plugins/mwiki/index.php/mricros:MainPage>). Other figures were made using the Canlab Coretools toolbox (<https://github.com/canlab/CanlabCore>) and custom code.

**Psychophysiological interaction (PPI) analysis.** The PPI analysis was performed using the CONN FC toolbox (<http://web.mit.edu/swg/software.htm>). The GLM comprised regressors for attention condition (full/diverted), confidence (confident/guess), and expectation-response congruence (congruent/neutral/incongruent). Nuisance regressors were identical to those used in the GLM on BOLD, with the addition of scrubbing parameters, which exclude outlier volumes that bias connectivity estimates. Again, the signal was not bandpass filtered, but instead the mean WM drift was entered as a nuisance regressor. Otherwise, default preprocessing steps were taken: the data were denoised by regressing out signal from WM, from CSF and from each individual condition, plus signal associated with all nuisance regressors. The PPI was run on univariate regression weights to identify effective connectivity between a functionally defined seed (rIFG) and remaining voxels. These weights were examined in a second-level model, which used an uncorrected peak threshold of  $p < 0.005$  and FDR cluster corrected threshold of  $p < 0.05$ .

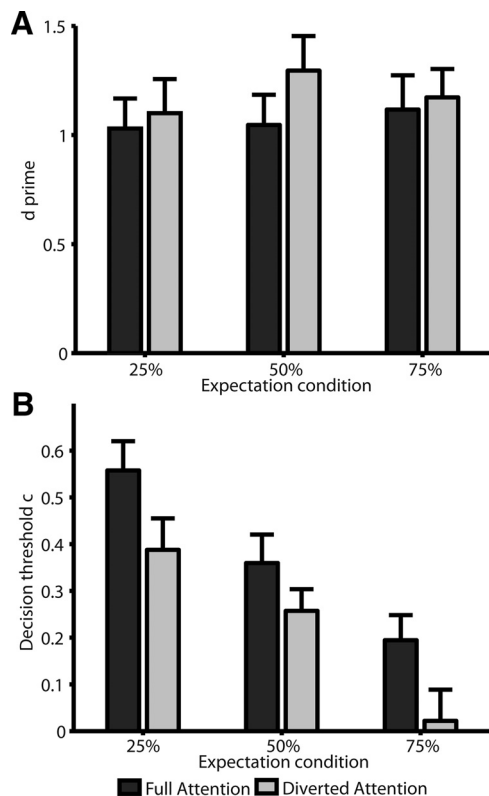
**Voxel-based morphometry.** T1-weighted structural scans were reoriented to the anterior commissure and segmented into GM, WM, and CSF. These were normalized to MNI space using DARTEL with SPM defaults and a Gaussian smoothing kernel of 8 mm FWHM (Ashburner and Friston, 2000). Modulated WM and GM images were separately compared across participants in a multiple regression with age and total intracranial volume (GM + WM + CSF) as nuisance regressors. Gender was not included because this resulted in multicollinearity between regressors (older participants were more likely to be male). Clusters reported as significantly correlating with behavior survived voxelwise FWE correction.

## Results

### Expectations liberalize decision thresholds and attention increases contrast sensitivity

Our first analyses confirmed the efficacy of our paradigm. To equate difficulty across attention conditions and participants, adaptive psychophysical staircases identified the stimulus contrast required for 79% accuracy on the Gabor detection task (see Staircases). Comparing the acquired contrasts in the full ( $4.34 \pm 3.50\%$ , mean  $\pm$  SD) and diverted ( $5.69 \pm 3.79\%$ ) attention conditions revealed that contrast thresholds were significantly lower under full than diverted attention ( $t_{(19)} = 2.95$ ,  $p = 0.014$ , 95% CI [0.50%, 2.31%],  $d_z = 0.70$ ), indicating that our paradigm successfully manipulated attention.

To ensure that our staircase procedure successfully equated detection sensitivity  $d'$  across conditions, we ran a within-subjects Attention (full, diverted)  $\times$  Expectation (25%, 50%, 75%) ANOVA. This revealed no significant difference between  $d'$  under full ( $1.06 \pm 0.14$ ) and diverted ( $1.21 \pm 0.20$ ) attention conditions ( $F_{(1,18)} = 0.34$ ,  $p = 0.569$ ,  $\eta_p^2 = 0.02$ ) (Fig. 2A), and was corroborated by a Bayesian repeated-measures ANOVA of the same design that revealed moderate evidence for the null hypothesis (BF = 0.240). There was no significant effect of Expectation on  $d'$  ( $F_{(2,36)} = 0.70$ ,  $p = 0.505$ ,  $\eta_p^2 = 0.04$ , BF = 0.07; strong evidence for the null) and no significant interaction term ( $F_{(2,36)} = 0.76$ ,  $p = 0.476$ ,  $\eta_p^2 = 0.04$ , BF = 0.016; strong evidence for the null). Our staircases therefore successfully equated  $d'$ .



**Figure 2.** Behavioral effects of expectation and attention on objective and subjective decision-making. **A**, Detection sensitivity  $d'$  as a function of expectation and attention condition. No significant differences were found. **B**, Decision threshold  $c$  as a function of expectation and attention condition. Independently of attention, bias toward reporting “yes” (lower values of  $c$ ) increased with the prior probability of Gabor presence. Error bars indicate within-subjects within-subject SEM.

To determine whether we had successfully manipulated priors, we compared signal detection theoretic decision thresholds ( $c$ , see Materials and Methods) across levels of expectation (de Lange et al., 2013; Morales et al., 2015; Sherman et al., 2015). As the expectation of Gabor presence over absence increases, decision threshold should become increasingly biased toward “yes” responses (i.e., liberalized, indicated by lower values of  $c$ ). This was confirmed in a within-subjects Attention (full, diverted)  $\times$  Expectation (25%, 50%, 75%) ANOVA ( $F_{(1,65,29.72)} = 18.10, p < 0.001, \eta_p^2 = 0.50$ ). LSD *post hoc* tests revealed a greater bias toward reporting “yes” in the 50% (neutral) than the 25% (expect absent) condition ( $p = 0.010, d_z = 1.15$ ), and in the 75% (expect present) than the 50% (neutral) condition ( $p < 0.001, d_z = 1.39$ ) (Fig. 2B). We found no evidence for attentional effects on decision threshold ( $F_{(1,18)} = 3.38, p = 0.083, \eta_p^2 = 0.16$ ) and no Expectation  $\times$  Attention interaction ( $F_{(2,36)} = 0.37, p = 0.693, \eta_p^2 = 0.020$ ).

Summarizing these results, our design successfully independently manipulated attention and expectation, while keeping detection sensitivity constant across conditions.

### Expectations increase confidence

We have previously shown that subjective confidence increases when perceptual decisions are congruent with prior expectations (Sherman et al., 2015), and on this basis, hypothesized that confidence would relate to prediction error signals. To determine whether we had replicated this behavioral result, we compared confidence for expectation-congruent and expectation-inc-

ongruent perceptual decisions. Congruent responses are “yes” reports in the 75% (expect present) condition and “no” reports in the 25% (expect absent) condition. The reverse applies for incongruent responses. An effect of congruence is therefore demonstrated by an interaction between expectation and report, whereby when participants report “yes,” confidence increases with increasing prior probability of target presence, and when participants report “no” confidence increases with decreasing probability of target presence.

The percentage of high confidence trials were subjected to an Attention (full, diverted)  $\times$  Accuracy (correct, incorrect)  $\times$  Expectation (25%, 50%, 75%)  $\times$  Report (yes, no) repeated-measures ANOVA. Participants appropriately showed higher confidence in correct than in incorrect detection judgments ( $F_{(1,18)} = 54.583, p < 0.001, \eta^2 = 0.752$ ). Confidence was subject to several other effects. Overall, confidence was higher when reporting absence. This held for incorrect but not correct judgments, and under full but not diverted attention. Attention also interacted with expectation such that under attention confidence was higher when the prior was informative (the 25% and 75% conditions), but under diverted attention confidence was higher in the absence of an informative prior (50% condition). Statistics are presented in Table 1.

Crucially, congruent reports were associated with higher confidence than incongruent reports, as shown by a significant Expectation  $\times$  Report interaction ( $F_{(2,36)} = 15.535, p < 0.001, \eta^2 = 0.463$ ) (Fig. 3). Here, confidence for “yes” responses was higher in the 75% (congruent) than the 25% (incongruent) condition ( $t_{(18)} = 2.51, p = 0.021$ ), whereas confidence for “no” responses was higher in the 25% (congruent) than the 75% (incongruent) condition ( $t_{(18)} = 3.83, p = 0.001$ ). The three-way interaction with Attention was not significant ( $p = 0.938$ ). Confidence therefore increased when the reported percept was congruent with prior expectations, independently of attentional resource. In other words, perceptual decisions that were consistent with prior beliefs were associated with higher confidence. This crucial behavioral result motivated our investigation into the neural correlates of predictability effects on subjective confidence.

### Two forms of congruency

To unravel the neural correlates of predictive influences on confidence, we first needed to identify brain regions sensitive to perceptual expectations. We predicted, based on previous work, that areas sensitive to perceptual expectations would exhibit an increased BOLD amplitude for trials on which expectations were violated (Egner et al., 2010; Kok et al., 2012a; Jiang et al., 2013; Kouider et al., 2015; St. John-Saaltink et al., 2015).

The experimental design used near-threshold stimuli, leading to potential dissociations between stimulus presentation and perceptual contents. Accordingly, expectations can be violated in two ways. First, stimulus presentation can be unexpected. These violations are under the control of the experimenter. We refer to the neural correlate of this expectancy violation  $PE_{\text{STIMULUS}}$ . Second, the perceptual report (i.e., perceptual content) can be incongruent with that expected. These violations are not under the control of the experimenter. Rather, they are a function of the participant’s sensory representation. We refer to the neural correlate of this expectancy violation  $PE_{\text{REPORT}}$ .  $PE_{\text{STIMULUS}}$  is most often observed at lower levels of the perceptual hierarchy (Kok et al., 2012a; Chennu et al., 2013; Jiang et al., 2013), whereas the decision-related  $PE_{\text{REPORT}}$  signals have been reported in both visual cortex (Pajani et al., 2015) and higher-level, decision-related areas (Bubic et al., 2009).

**Table 1. Significant results from four-way repeated-measures ANOVA on behavioral confidence**

	ANOVA			Follow-up contrast	Post-hoc <i>t</i> test			
	<i>F</i>	<i>p</i>	$\eta^2$		<i>t</i>	<i>p</i>	<i>M</i> <sub>diff</sub> (%)	<i>SE</i> <sub>diff</sub> (%)
Response	9.32	0.007	0.341	No-yes	0.83	0.416	12.71	4.15
Response × accuracy	71.34	<0.001	0.799	No-yes (correct)	7.85	<0.001	−4.40	5.28
				No-yes (incorrect)	5.26	<0.001	30.93	3.94
				No-yes (full)	0.52	0.607	22.20	4.22
Response × attention	15.41	0.001	0.461	No-yes (diverted)	3.19	0.005	2.87	5.49
				Quadratic trend 25% + 75% > 50% (full)	2.27	0.036	5.12	1.60
Expectation × attention	8.65	0.001	0.325	Quadratic trend 50% > 25% + 75% (diverted)			5.86	2.58

### Representation of PE<sub>STIMULUS</sub> in visual cortex

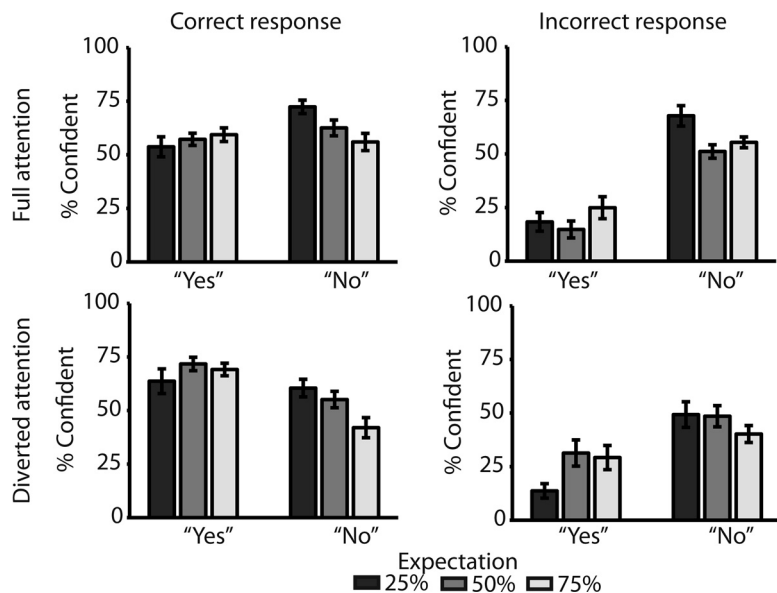
In our first analysis, we searched for regions that are sensitive to discrepancies between expectation and stimulus presentation over whole brain. To do this, we computed the contrast unexpected stimulus presentation > expected stimulus presentation. Gabor presence is expected in the 75% condition but unexpected in the 25% condition, whereas Gabor absence is expected in the 25% condition but unexpected in the 75% condition. Our analysis identified one PE<sub>STIMULUS</sub>-sensitive area in contralateral occipital cortex (V1-V3, BA18, peak MNI  $x = -12$ ,  $y = -80$ ,  $z = 22$ ,  $Z_{\text{peak}} = 4.09$ ,  $0.66 \text{ cm}^3$ , cluster  $p_{\text{FDR}} = 0.350$ ,  $p_{\text{uncorr}} = 0.023$ ) and one on the ipsilateral side (V1-V3, BA18, peak MNI  $x = 8$ ,  $y = -80$ ,  $z = 18$ ,  $Z_{\text{peak}} = 3.99$ ,  $1.01 \text{ cm}^3$ , cluster  $p_{\text{FDR}} = 0.205$ ,  $p_{\text{uncorr}} = 0.007$ ). Neither of these clusters survived cluster-level correction, so they will not be considered beyond this point. They are presented simply to show consistency with previous studies, in which statistical power was improved by constraining the analysis with functional localizers (Smith and Muckli, 2010; Kok et al., 2012a, b; Larsson and Smith, 2012; Jiang et al., 2013).

The whole-brain contrast PE<sub>STIMULUS</sub>, attended > PE<sub>STIMULUS</sub>, unattended yielded no significant or marginally significant clusters, indicating no evidence for a PE<sub>STIMULUS</sub> × attention interaction.

Using a peak threshold of  $p < 0.005$ , both of these analyses were replicated under our control model, which included reaction speed and Gabor contrast as parametric modulators (unexpected > expected, contralateral:  $p_{\text{FDR}} = 0.446$ ,  $p_{\text{uncorr}} = 0.014$ , ipsilateral:  $p_{\text{FDR}} = 0.446$ ,  $p_{\text{uncorr}} = 0.011$ ).

### Regions representing PE<sub>REPORT</sub>

Next, we searched for regions whose BOLD response reflects the discrepancy between expectation and perceptual report. Expectation-congruent reports are “yes” responses in the 75% (expect present) condition and “no” responses in the 25% (expect absent) condition. The reverse applies for expectation-incongruent reports. These definitions differ from those in the previous analysis because they consider perceptual report instead of stimulus presence or absence. In turn, this analysis searches for regions sensitive to unexpected perceptual content.



**Figure 3.** Confidence as a function of attention, accuracy, report, and expectation. Each panel represents confidence as a function of expectation and yes/no report. Panels in the left column represent confidence for correct responses. Panels in the right column represent confidence for incorrect responses. Panels in the top row represent confidence under full attention. Panels in the bottom row represent confidence under divided attention. Independently of accuracy (column) and attention (row), confidence increases with expectation-response congruence. Within each panel, confidence for “yes” responses increases with increasing prior probability of presence (i.e., increasing congruence), and for “no” responses it decreases with increasing prior probability of presence (i.e., decreasing congruence). Error bars indicate within-subjects SEM.

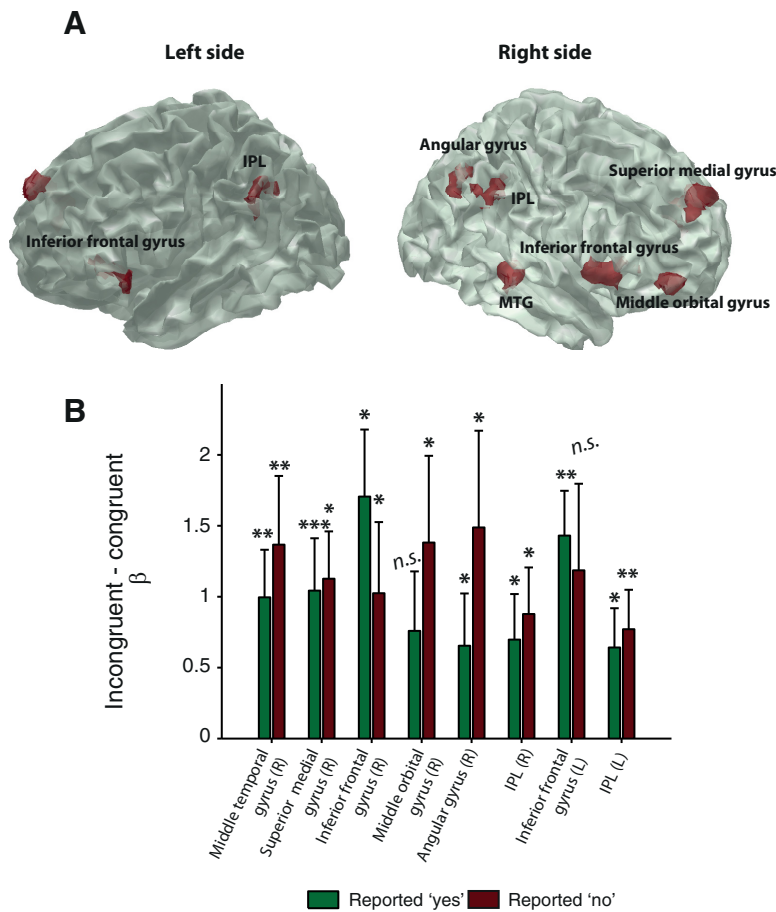
The contrast expectation-incongruent report > expectation-congruent report was computed over whole brain. This revealed eight significant clusters distributed throughout the cortex (Fig. 4A; Table 2). We found no significant clusters for the reverse contrast, even with a more liberal peak threshold of  $p < 0.005$  uncorrected.

Regions exhibiting a PE<sub>REPORT</sub> pattern should show a larger BOLD response to incongruent than congruent responses, regardless of whether the response was “yes” or “no” (Kok et al., 2012a). We checked whether this held in these putative PE<sub>REPORT</sub> regions by extracting median regression coefficients as a function of attention, expectation and report, and subjecting them to separate repeated-measures ANOVAs.

Results are depicted in Figure 4B, and statistics are presented in Table 3. All regions exhibited a significant PE<sub>REPORT</sub> response for both “yes” and “no” judgments, except middle orbital gyrus and left inferior frontal gyrus. Accordingly, these clusters were excluded from our set of PE<sub>REPORT</sub> regions.

Under our control model (in which regressors were parametrically modulated by Gabor contrast and reaction speed, see Materials and Methods), all significant results here were





**Figure 4.** Report prediction error. **A**, Results of contrast incongruent response > congruent response over whole brain. Only clusters surviving FDR cluster correction are shown. **B**,  $PE_{REPORT}$  (incongruent – congruent), by region and perceptual report.  $PE_{REPORT}$  for “yes”/“no” reports are presented in green/red, respectively. BOLD has been averaged over levels of attention. \* $p < 0.05$ ,  $PE_{REPORT}$  significantly different from zero. \*\* $p < 0.01$ ,  $PE_{REPORT}$  significantly different from zero. \*\*\* $p < 0.001$ ,  $PE_{REPORT}$  significantly different from zero. Error bars indicate SEM.

**Table 2. Results of whole-brain analysis expectation-incongruent report > expectation-congruent report<sup>a</sup>**

Region	BA	Side	Volume (cm <sup>3</sup> )	Peak Z	$p_{FDR}$	Peak MNI		
						x	y	z
MTG	21	R	2.29	4.78	0.007	54	−30	−2
SMG	9/10	R	4.15	4.54	<0.001	12	58	32
IFG	47/48	R	2.70	4.45	0.004	56	12	−2
MOG	47/46	R	2.08	4.33	0.009	40	50	−6
AG	39	R	1.21	3.95	0.044	46	−64	36
IPL	40	R	1.21	3.91	0.044	58	−40	40
IFG	47	L	1.90	3.79	0.012	−38	26	−4
IPL	40/48	L	1.60	3.75	0.021	−54	−46	34

<sup>a</sup>MTG, Middle temporal gyrus; SMG, superior medial gyrus; MOG, middle orbital gyrus; AG, angular gyrus; IPL, inferior parietal lobule.

replicated, at least at marginal significance (rIFG, our critical region,  $p_{FDR} = 0.044$ ). Results were fully replicated when using a peak threshold of  $p < 0.005$ . This means that  $PE_{REPORT}$  responses do not reflect reaction speed or Gabor contrast/detection sensitivity.

In summary, we have identified six regions signaling  $PE_{REPORT}$  as follows: right middle temporal gyrus, right superior medial gyrus, rIFG, right angular gyrus, and bilateral inferior parietal lobule. These regions are sensitive to the discrepancy between perceptual expectations and perceptual choice (i.e., to unexpected perceptual inferences).

**High confidence is associated with an attenuated  $PE_{REPORT}$  response in rIFG**

Our key hypothesis was that high confidence would be associated with low  $PE_{REPORT}$ . However, confidence can be also influenced by attention (Rahnev et al., 2011) and tracks accuracy (Dienes, 2008; Pleskac and Busemeyer, 2010). To test whether any  $PE_{REPORT}$  region represented confidence after controlling for these potential confounds, median regression weights from each  $PE_{REPORT}$  region were extracted as a function of confidence, attention, and decision accuracy. These regression coefficients were then subjected to separate Bayesian repeated-measures ANOVAs. We were looking for regions whose BOLD response (in these regions, representing  $PE_{REPORT}$ ) differs with confidence. We could not test for a  $PE_{REPORT} \times$  Confidence interaction: when the participant reports low confidence, they have signaled their yes/no decision as unreliable; that is, their perception of Gabor presence or absence does not necessarily correspond to their report. Because  $PE_{REPORT}$  is a function of both Expectation and Report, this variable will also be unreliable when the participant has reported low confidence.

Only one region exhibited a BOLD response (i.e.,  $PE_{REPORT}$  amplitude) that differed as a function of subjective confidence: rIFG. Here, supporting our hypothesis, BOLD amplitude was higher for guess than confident reports (Fig. 5A). Crucially, the analysis revealed substantially more evidence for modeling rIFG

BOLD as a function of confidence alone ( $BF = 13.620$ ) than as a function of just accuracy ( $BF = 0.877$ ), just attention ( $BF = 0.711$ ), or as a combination of confidence and any other factors ( $BF = 0.003–2.069$ ; for summary of results from all ROIs, see Table 4). A frequentist ANOVA gave the same result: a significantly higher BOLD amplitude for guess than confident responses ( $F_{(1,18)} = 6.04$ ,  $p = 0.024$ ,  $\eta^2 = 0.251$ , 95% CI [0.10, 1.28]). These results are shown in Figure 5B.

Does rIFG sensitivity to confidence indeed reflect an effect of expectation? If so, the effect of expectation on confidence (as shown in Fig. 3) should be correlated with IFG  $PE_{REPORT}$  amplitude. We tested this with an across-subject brain-behavior correlation. Our behavioral variable was the effect of expectations on confidence, denoted  $\Delta\text{Confidence} = \text{Confidence}_{\text{congruent}} - \text{Confidence}_{\text{incongruent}}$ . Because  $PE_{REPORT}$  cannot be meaningfully computed for guess responses (see above), it was computed from confident responses only.

Correlating these two variables revealed a significant negative correlation ( $\rho = -0.512$ ,  $p = 0.027$ ) (Fig. 5C): smaller  $PE_{REPORT}$  amplitude in rIFG was associated with larger increases in confidence for expectation-congruent perceptual decisions. In turn, this result confirms our finding that high confidence is associated with low  $PE_{REPORT}$  in rIFG, and implicates rIFG in the mechanism by which prior expectations increase confidence.



**Table 3. Effect of expectation, separately for “yes” and “no” reports<sup>a</sup>**

Region	Reported “no”			Reported “yes”			PE <sub>REPORT</sub> ?
	<i>F</i>	<i>p</i>	$\eta^2$	<i>F</i>	<i>p</i>	$\eta^2$	
MTG	8.82	0.008	0.329	5.83	0.006	0.245	Yes
SMG	8.10	0.001	0.310	4.46	0.014	0.213	Yes
rIFG	4.70	0.015	0.207	3.45	0.041	0.162	Yes
MOG	1.95	0.157	0.098	3.42	0.044	0.160	No
AG	3.52	0.040	0.164	4.07	0.025	0.185	Yes
rIPL	4.71	0.044	0.207	7.17	0.015	0.285	Yes
lIFG	5.62	0.008	0.238	2.87	0.070	0.137	No
lIPL	5.39	0.032	0.230	6.04	0.005	0.251	Yes

<sup>a</sup>Both effects should be significant for the region to be deemed a PE<sub>REPORT</sub> region. MTG, Middle temporal gyrus; SMG, superior medial gyrus; MOG, middle orbital gyrus; AG, angular gyrus; rIPL, right inferior parietal lobule; lIFG, left inferior frontal gyrus; lIPL, left inferior parietal lobule.

To ensure that these differences were not driven by differences in reaction speed or Gabor contrast, we extracted data from the rIFG cluster revealed by our control GLM. This revealed that, even after controlling for these possible confounds, rIFG BOLD was significantly higher for guess than confident responses ( $t_{(18)} = 2.21$ ,  $p = 0.041$ ,  $d_z = 0.44$ ). The significant brain-behavior correlation was also replicated ( $\rho = -0.510$ ,  $p = 0.027$ ).

Together, these analyses reveal that subjective confidence is reliably associated with PE<sub>REPORT</sub> in right IFG even after controlling for attention, Gabor contrast, decision accuracy, and reaction speed.

### Sources of priors and sensory signals for confidence

Thus far, we have shown that lower PE<sub>REPORT</sub> amplitude in rIFG, reflecting the extent to which perceptual inferences are unexpected, are associated with higher confidence. Assuming a model in which decision confidence is a weighted function of top-down expectations and “bottom-up” sensory signals (or decision evidence), we asked whether we could identify sources of these variables. To do this, we ran a seed-to-voxel psychophysiological interaction analysis (PPI), with rIFG as a functionally defined seed.

We were interested in regions communicating predictive information and therefore searched for clusters exhibiting different degrees of FC with rIFG for congruent and incongruent reports. We reasoned that, although confidence should be a function of both sensory signals and expectations, there would be individual differences in how each component would be weighted, reflecting, for example, how reliable the expectation information is thought to be. Capitalizing on these individual differences, we reasoned that rIFG would show stronger FC with the expectation region in participants whose confidence was weighted more by expectation. By contrast, rIFG would show stronger FC with the source of sensory signals in participants whose confidence was only weakly shaped by expectation.

To test this hypothesis, we used a behavioral covariate of interest – the influence of expectations on confidence. This behavioral variable was defined as  $\Delta\text{Confidence} = \text{Confidence}_{\text{Incongruent}} - \text{Confidence}_{\text{Congruent}}$ , and is the same as the behavioral variable in Figure 5C. Higher values signify that expectations exerted a stronger effect on confidence.

Sources of predictive information for confidence were identified by computing the contrast incongruent  $\neq$  congruent, with  $\Delta\text{Confidence}$  as a between-subjects covariate of interest. That is, we searched for brain connectivity-behavior relationships.

As shown in Figure 6A, the PPI analysis revealed three significant clusters. The more expectations shaped confidence (higher

$\Delta\text{Confidence}$ ), the more that congruence was associated with FC between rIFG and two clusters: one in left orbitofrontal cortex (IOFC) ( $\Delta\text{Confidence} \times \text{congruent} > \text{incongruent}$ ; peak MNI  $x = -36$ ,  $y = 38$ ,  $z = -18$ ,  $3.40 \text{ cm}^3$ , cluster  $p_{\text{FDR}} = 0.008$ ) and one in right OFC ( $\Delta\text{Confidence} \times \text{congruent} > \text{incongruent}$ ; peak MNI  $x = 10$ ,  $y = 26$ ,  $z = -18$ ,  $2.50 \text{ cm}^3$ , cluster  $p_{\text{FDR}} = 0.024$ ).

By contrast, the less expectations shaped confidence (lower  $\Delta\text{Confidence}$ ), the more that congruence was associated with FC between rIFG and intracalcarine sulcus ( $\Delta\text{Confidence} \times \text{congruent} > \text{incongruent}$ ; peak MNI  $x = 6$ ,  $y = -58$ ,  $z = 12$ ,  $3.86 \text{ cm}^3$ , cluster  $p_{\text{FDR}} = 0.004$ ). Thus, intracalcarine sulcus and bilateral OFCs exhibited a push-pull relationship, with the dominant region predicted by  $\Delta\text{Confidence}$ .

Although the balance of FC between these regions was determined by  $\Delta\text{Confidence}$ , FC between these rIFG and these regions was present independently of  $\Delta\text{Confidence}$ . Specifically, FC between rIFG and OFC was greater than zero on congruent trials (IOFC,  $p < 0.001$ ; rOFC,  $p = 0.053$ ), whereas intracalcarine sulcus-rIFG FC was significantly greater than zero on incongruent trials ( $p = 0.022$ ).

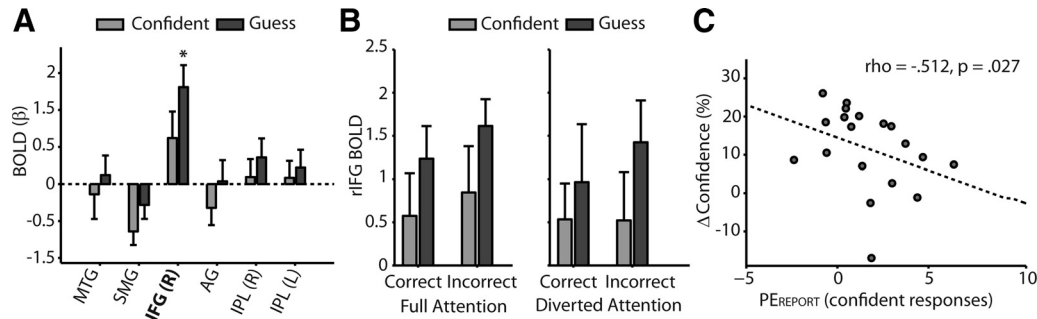
Because OFC was primarily associated with congruent responses, we reasoned that FC with these regions might reflect the communication of perceptual priors. Consistent with this, we found a main effect of expectation condition on IOFC BOLD ( $F_{(1.64, 29.59)} = 3.61$ ,  $p = 0.047$ ,  $\eta_p^2 = 0.167$  (Fig. 5B). Here, BOLD exhibited a ‘U’-shaped relationship with expectation ( $p = 0.036$ ), consistent with the representation of prior information: BOLD was higher when there was an informative prior (the 25% and 75% conditions) than when the prior was flat (the 50% condition). In rOFC, this pattern was exhibited under full ( $F_{(2, 36)} = 3.80$ ,  $p = 0.032$ ,  $\eta_p^2 = 0.174$ ) but not diverted ( $F_{(2, 36)} = 0.87$ ,  $p = 0.426$ ,  $\eta_p^2 = 0.046$ ) attention (interaction  $p = 0.030$ , Fig. 6C). Interestingly, in rOFC, there was also a significant attention  $\times$  confidence interaction ( $F_{(1, 18)} = 7.84$ ,  $p = 0.012$ ,  $\eta_p^2 = 0.303$ ) (Fig. 6D), such that attention reversed the BOLD response to confident versus guess responses.

These results are consistent with the interpretation of bilateral OFC communicating prior information. Whereas IOFC represented prior information independently of attention, rOFC did this only under full attention. Moreover, the attention by confidence interaction under rOFC BOLD suggests that this region may additionally represent the degree of (reverse) uncertainty associated with attentional state.

Next, we asked whether intracalcarine sulcus represented prediction error signals. PE<sub>REPORT</sub> is demonstrated in an expectation by report interaction, whereas PE<sub>STIMULUS</sub> is demonstrated in an expectation by stimulus interaction. However, neither was found (all  $p > 0.288$ ). Rather, the BOLD response here was marginally higher for stimulus present than absent trials,  $F_{(1, 18)} = 3.81$ ,  $p = 0.067$ ,  $\eta_p^2 = 0.175$  (Fig. 6E).

One might wonder whether bilateral OFC directly signals priors to occipital lobe, or vice versa for sensory signals. This was not the case. Rerunning the PPI analysis in the same way, but with either OFC cluster as our seed, revealed no significant or marginally significant connectivity with intracalcarine sulcus. Similarly, running the analysis setting intracalcarine sulcus as the seed revealed no significant or marginally significant connectivity with either OFC cluster.

Together, these results show that the integration of expectations into confidence judgments in rIFG recruits an occipitofrontal network that represents top-down influences of attention and



**Figure 5.** The relationship between confidence and  $PE_{REPORT}$ . **A**, BOLD as a function of confidence in each  $PE_{REPORT}$  region. BOLD is significantly higher for guess than confident responses in rIFG only. **B**, rIFG BOLD as a function of attention, accuracy, and confidence. BOLD is higher for guess than confidence responses independently of attention and decision accuracy. **C**, Brain-behavior correlation.  $PE_{REPORT}$  in rIFG is negatively correlated with the behavioral increase in confidence by expectation ( $\Delta$ Confidence). \* $p < 0.05$ . Error bars indicate within-subject SEM.

**Table 4. Results of Bayesian confidence  $\times$  accuracy  $\times$  attention repeated-measures ANOVAs<sup>a</sup>**

Region	Bayes factors <sup>b</sup>				Null
	Confidence	Accuracy	Attention	Confidence + others	
MTG	0.14	0.68	4.96	0.01–2.26	1.00
SMG	0.85	1.79	1.96	0.03–2.67	1.11
rIFG	13.62	0.88	0.71	<0.01–2.07	3.96
AG	1.35	9.83	0.53	0.01–3.42	3.47
rIPL	2.64	0.84	2.42	<0.01–1.13	10.07
lIPL	1.74	5.52	1.84	<0.01–1.17	10.22

<sup>a</sup>MTG, Middle temporal gyrus; SMG, superior medial gyrus; AG, angular gyrus; rIPL, right inferior parietal lobe; lIPL, left inferior parietal lobe.

<sup>b</sup>Bayes factors correspond to the evidence for the listed model relative to the evidence for all other models.

expectation in OFC, and bottom-up sensory signals in intracalcarine sulcus.

### The contribution of OFC to confidence is predicted by WM density

Our connectivity analyses revealed that the integration of expectations into confidence judgments recruits OFC, representing top-down signals, and intracalcarine sulcus, representing bottom-up signals. The extent to which each region was recruited was predicted by individual differences in the extent to which expectations shaped behavioral confidence. The presence of these individual differences motivated an exploratory follow-up analysis that asked whether they are reflected in brain structure. More specifically, we considered whether the weighting of top-down predictions was a function of WM or GM density of the source region.

The BOLD response of OFC reflected an effect of perceptual expectations on objective decision. The behavioral correlate of this is therefore  $\Delta c = c_{25\%} - c_{75\%}$ , the extent to which perceptual expectations bias (yes/no) decision. We performed a whole-brain multiple regression analysis on WM density, with total intracranial volume and participant age as nuisance covariates, and with  $\Delta c$  as the regressor of interest. This analysis revealed that propensity to incorporate low-level priors into decision-making, as measured by  $\Delta c$ , was negatively correlated with rOFC WM density (Fig. 7; peak MNI  $x = 23$ ,  $y = 30$ ,  $z = -14$ ,  $11.51\text{cm}^3$ ,  $p_{\text{peak-FWE}} = 0.030$ ,  $Z = 5.08$ ). The same analysis for GM yielded no significant results.

This result suggests that the dependence of confidence on FC with the expectation source regions is reflected in anatomical indications of that connectivity: WM density in OFC was negatively predicted by its functional correlate.

### Discussion

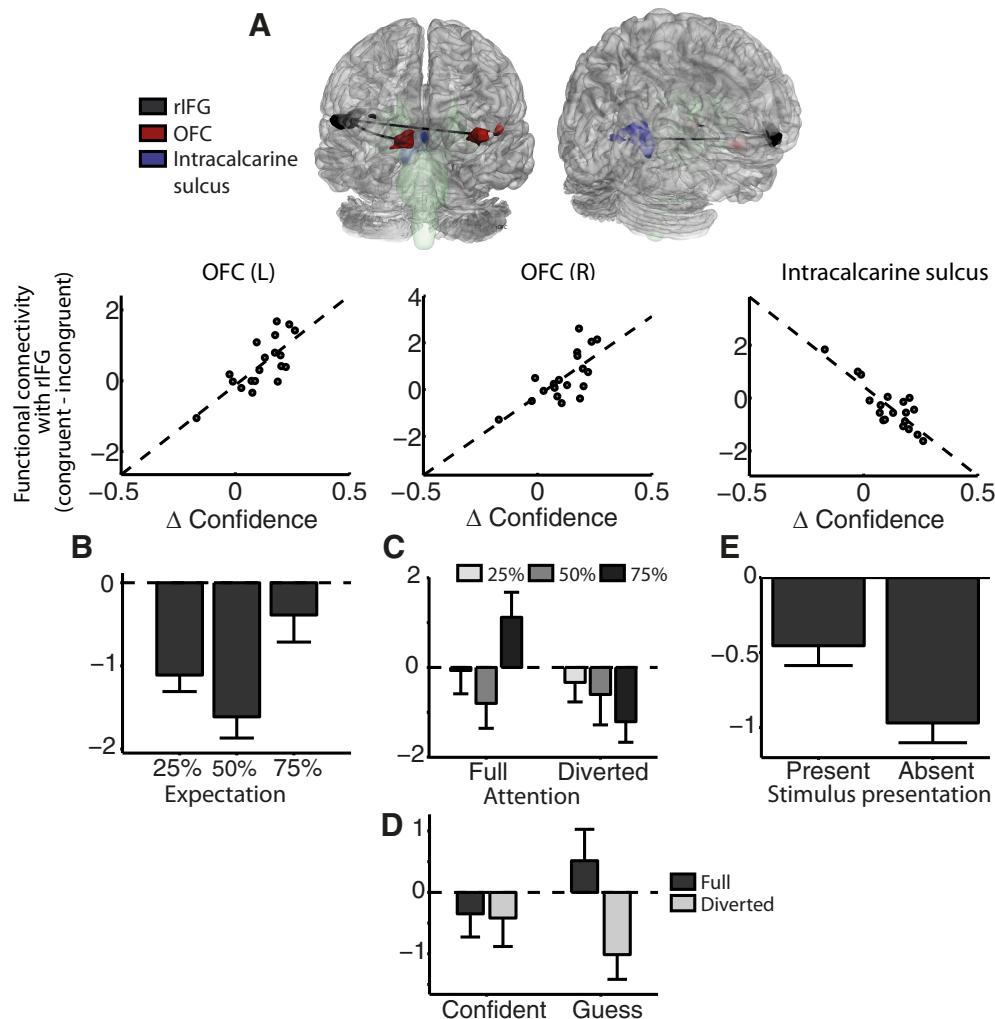
In the present paper, we have shown that perceptual confidence increases when decisions are supported by (or congruent with) prior expectations. Crucially, we show that the process of integrating this predictive information recruits rIFG.

Our data reveal that unexpected perceptual content is associated with heightened  $PE_{REPORT}$  (a mismatch response to expectation-incongruent perceptual decisions) in a distributed set of frontal, parietal, and temporal decision-related regions. Interestingly, this  $PE_{REPORT}$ -sensitive set resembles those implicated in other forms of “top-down” processing, such as modality-independent sensory change detection (Downar et al., 2000), response inhibition (Verbruggen and Logan, 2008; Criado and Boulinguez, 2013), and detection of behavioral salience (Downar et al., 2002).

Crucially, the contribution of top-down expectations to subjective confidence judgments was reflected in rIFG BOLD activity. Here, high confidence was associated with a lower  $PE_{REPORT}$  response. Furthermore, stronger behavioral effects of expectation on confidence were associated greater  $PE_{REPORT}$  attenuation in this region. Control analyses ruled out explanations in terms of attention, stimulus contrast, decision accuracy, or reaction speed. Our results therefore indicate a central role for rIFG in subjective perceptual decision making, in which the “mismatch” between internal templates and perceptual content is integrated into subjective confidence judgments. That is, these data indicate that perceptual decision confidence is shaped by a neural mismatch response that is perceptual in nature.

One might interpret  $PE_{REPORT}$  as arising from the preparation or execution of a response that is a priori likely to be incorrect. Recent evidence has shown that perceptual confidence is increased by response-effector congruency (Fleming et al., 2015). However, we find such an explanation for the present results unlikely for two reasons: (1) our paradigm did not experimentally induce response conflict; and (2) effects of confidence on rIFG BOLD were independent of decision accuracy. Rather, we suggest that  $PE_{REPORT}$  reflects perceptual “prediction error,” arising from the degree to which the reported percept is surprising.

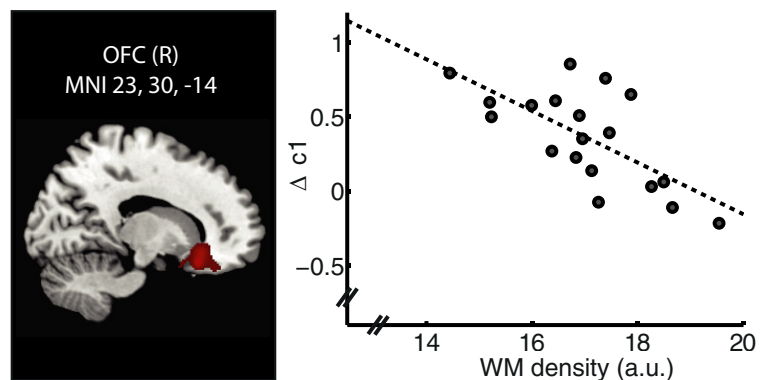
Our data suggest that expectations shape the decision signal on which confidence judgments are based (although we cannot and do not rule out an additional effect of expectations at the metacognitive level). Accordingly, our results are consistent with many models of choice confidence that permit expectations to shape decision evidence (for a thorough review, see Summerfield and de Lange, 2014). For example, in recent work, we have shown that the signal detection model can be reformulated in Bayesian terms. Here, the distributions of internal responses to stimulus presence and absence are



**Figure 6.** PPI analysis. **A**, Clusters that exhibit congruence-dependent FC with rIFG that is correlated with  $\Delta$ Confidence. The analysis revealed intracalcarine sulcus (blue) and bilateral OFC (red). The black cluster is our seed, rIFG. Black lines joining clusters indicate FC between regions. Scatterplots below represent the correlation between each region's congruence-dependent FC with rIFG and  $\Delta$ Confidence. **B**, Significant quadratic modulation of left OFC BOLD by expectation. **C**, Right OFC BOLD as a function of attention and expectation. There is a significant effect of expectation under full but not diverted attention. **D**, Right OFC BOLD as a function of attention and confidence. Attention reverses the effect of confidence on BOLD. **E**, Marginally lower intracalcarine sulcus BOLD on Gabor absent than on Gabor present trials. Error bars indicate within-subject SEM.

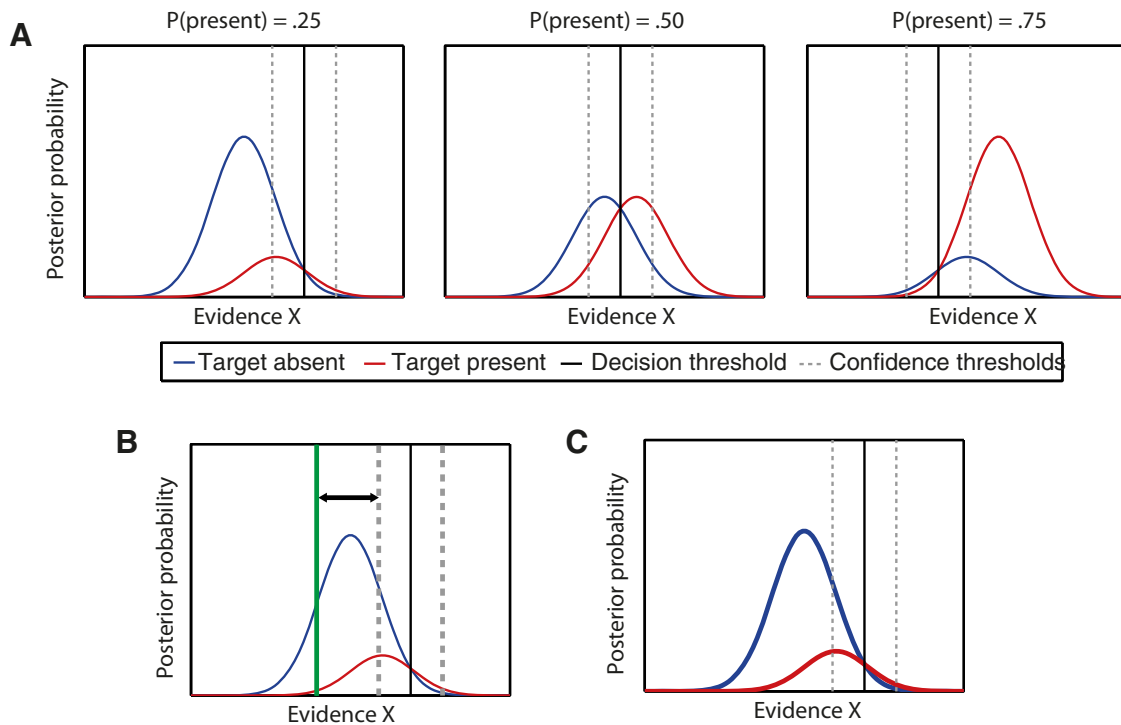
represented as posterior probability distributions, whereas decision and confidence thresholds are fixed according to a particular posterior odds ratio (Fig. 8A) (Sherman et al., 2015). Increasing the prior probability of target presence will change the distribution of internal responses, driving decision and confidence thresholds. The result of this is a greater propensity to report “yes” and to report “confident” when target presence expected. Crucially, these are not “undesirable” biases that ought to be eliminated. Rather, they reflect the underlying statistics of the task, even influenced by periodic fluctuations in occipital cortex (Sherman et al., 2016).

In the present study, one could argue that rIFG serves to compare this internal response to a confidence threshold (Fig. 8B). However, such an account would not be consistent with previous work, implicating only left IFG



**Figure 7.** Voxel-based morphometry results. WM density in right OFC negatively predicts the effect of expectation on perceptual decision.

in adjusting perceptual criteria, and right IFG in estimating task difficulty and uncertainty (Tops and Boksem, 2011; White et al., 2012; Reckless et al., 2014). However, this literature would support an account by which rIFG integrates prior



**Figure 8.** Bayesian signal detection model presented in Sherman et al. (2015) and possible roles for rIFG in this study. **A**, In the signal detection model presented in Sherman et al. (2015), the internal response to stimulus presence (red) or absence (blue) represents the posterior probability of that sensory event, given the evidence and the prior. Solid black line indicates the decision threshold, which is the posterior odds of presence to absence. The confidence thresholds (dashed gray lines) flank the decision threshold. Leftmost panel, Case where target absence is more likely than presence, and accordingly the peak of the stimulus absent distribution is higher. As the prior on presence increases (rightwards panels), the difference between these peaks changes accordingly. Similarly, while the decision threshold remains at the intersection of the two distributions, changes in the internal response drive the criteria to appear more liberal. This leads to apparent shifts in the accompanying confidence criteria, which are most liberal (have the largest area under the curve) when the congruent response is made. **B**, **C**, Alternative interpretations on the role of rIFG. **B**, rIFG may serve to compare the confidence threshold against the current evidence (green). **C**, rIFG may be involved in constructing the posterior probability distributions (bold).

beliefs into the internal response (Fig. 8C), forming the signal upon which confidence is constructed.

We found that the process of relating predictive information into confidence judgments recruited both intracalcarine sulcus and bilateral OFCs. Intracalcarine sulcus exhibited a marginally significant BOLD response to the stimulus, so we interpret connectivity with this region as reflecting the communication of sensory signals. We assume this effect was weak because it was localized to a large cluster while reflecting the neural response to a small stimulus in retinotopically organized space.

We further found that both right and left OFC represented top-down prior information, consistent with previous work (Schoenbaum and Roesch, 2005; Wallis, 2007; Trapp and Bar, 2015). For left OFC, this was independent of attention, but right OFC was sensitive to attentional state: representation of the prior required attention. Furthermore, the effect of confidence on rOFC BOLD reversed under diverted attention, possibly indicating that rOFC represents the uncertainty associated with inattention.

Together, our results suggest that the construction of confidence recruits intracalcarine sulcus, representing stimulus-driven signals, OFC, representing top-down signals, and rIFG, which maps the discrepancy between expectation and perceptual choice to subjective confidence. OFC has been repeatedly been shown to reflect reward expectations and beliefs (Kepecs et al., 2008; Kim et al., 2011; De Martino et al., 2013; Lebreton et al., 2015); however, here we place OFC belief representations within a larger hierarchical structure for perceptual processing, perhaps generating predictions (Stalnaker et al., 2015; Trapp and Bar, 2015) that constrain perceptual confidence.

Importantly, our PPI analysis cannot determine the directionality of functional connections in this network. While the functional data are consistent with a model in which rIFG receives top-down and bottom-up inputs, on an alternative view bottom-up signals may be passed to rIFG, which passes a transformation of  $PE_{\text{REPORT}}$  into confidence to rOFC. Under this account, rIFG would construct the initial confidence representation, whereas OFC would transform the confidence estimate represented in rIFG into a reportable judgment, based on the mismatch between the estimate, expectations, and potentially, attentional state (Lebreton et al., 2015). A third possibility is that rIFG is recruited for the estimation of decision uncertainty, not confidence, and top-down processes from bilateral OFC are recruited more when this uncertainty is high. Therefore, top-down signals from rOFC would not shape rIFG representations of confidence but rather result from uncertain (i.e., variable or noisy) representations. Further studies will be needed sequence the involvement of these regions and disambiguate these possibilities.

Our results are readily interpretable from Bayesian brain perspectives (Lee, 2002; Yuille and Kersten, 2006; Friston, 2009; Clark, 2013). These propose that perceptual inference is a weighted integration of sensory evidence and prior beliefs about the cause of the sensation, such that the perceptual report corresponds to the belief with greatest posterior probability. The posterior probability increases as the correspondence between prior and sensory signal increases. Therefore, inference is deemed “successful,” and so should be associated with high confidence, when “prediction error” is low (Meyniel et al., 2015b), as we saw here. Neuronal representations of prediction errors are well established in the reward domain (Nakahara et al., 2004; Bayer and Glimcher, 2005), but in the perceptual domain evidence remains



restricted to BOLD correlates, such as  $PE_{\text{REPORT}}$ . From Bayesian perspectives, the finding that  $PE_{\text{REPORT}}$  amplitude in rIFG was lower for confident responses would point to the representation or construction of the posterior belief in this region; and indeed, rIFG has been found to encode the decision variable, both in Bayesian form (the posterior) (d'Acremont et al., 2013) and as decision evidence (Hebart et al., 2016).

Previous work has implicated rIFG in the representation of expectation violation in a range of modalities, from speech perception (Clos et al., 2014) to the auditory (Garrido et al., 2009), visual (Bubic et al., 2009), and tactile (Allen et al., 2016) domains. It has also been shown that rIFG represents subjective uncertainty (Fleck et al., 2006; Fleming and Dolan, 2012). However, to our knowledge, these functions of rIFG have not been related to each other before. rIFG is also frequently implicated in the seemingly disparate task of detecting or resolving response conflict (Casey et al., 2000; Hampshire et al., 2010), and is a key component of the response inhibition network (Verbruggen and Logan, 2008; Criaud and Boulinguez, 2013). This raises the intriguing possibility of a functional overlap between resolution of response conflict and the formation of confidence.

These roles could be unified by considering rIFG as the region in which the posterior belief on sensory causes is computed (at least for perceptual tasks) because the posterior affords a hypothesis space for adaptive, plausible actions (Mansouri et al., 2009). Consistent with this view, rIFG is recruited when appropriately acting on perceptual choices (Suzuki and Gottlieb, 2013), computing behavioral significance (Sakagami and Pan, 2007), and computing action-outcome likelihoods (Morris et al., 2014). Anatomical considerations also support such a view because the rIFG is directly connected with regions relevant for both cognitive and motor control (Petrides and Pandya, 2002). Right IFG may even compute high-level, abstracted posteriors that integrate high-level beliefs about decision outcomes because the rIFG BOLD response to erroneous decisions is associated with both the valence of the decision outcome, and the optimism of the participant ("self-belief") (Sharot et al., 2011). We leave open for future research the question of whether and how rIFG relates perceptual confidence to action outcomes.

In conclusion, we have shown that top-down expectations are integrated into decision confidence, and we have shown that this occurs in a functional network consisting of rIFG, OFC, and intracalcarine sulcus. Here, top-down perceptual expectations and bottom-up sensory inputs are integrated into a subjective sense of perceptual confidence. Together, our data reveal a crucial role of top-down influences in the mechanism by which perceptual decisions become available for conscious report.

## References

- Allen M, Fardo F, Dietz MJ, Hillebrandt H, Friston KJ, Rees G, Roepstorff A (2016) Anterior insula coordinates hierarchical processing of tactile mismatch responses. *Neuroimage* 127:34–43. [CrossRef Medline](#)
- Ashburner J, Friston KJ (2000) Voxel-based morphometry: the methods. *Neuroimage* 11:805–821. [CrossRef Medline](#)
- Bar M (2007) The proactive brain: using analogies and associations to generate predictions. *Trends Cogn Sci* 11:280–289. [CrossRef Medline](#)
- Bauer M, Stenner MP, Friston KJ, Dolan RJ (2014) Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. *J Neurosci* 34:16117–16125. [CrossRef Medline](#)
- Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47:129–141. [CrossRef Medline](#)
- Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Res* 49:1154–1165. [CrossRef Medline](#)
- Bubic A, von Cramon DY, Jacobsen T, Schröger E, Schubotz RI (2009) Violation of expectation: neural correlates reflect bases of prediction. *J Cogn Neurosci* 21:155–168. [CrossRef Medline](#)
- Bülthoff I, Bülthoff H, Sinha P (1998) Top-down influences on stereoscopic depth-perception. *Nat Neurosci* 1:254–257. [CrossRef Medline](#)
- Casey BJ, Thomas KM, Welsh TF, Badgaiyan RD, Eccard CH, Jennings JR, Crone EA (2000) Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proc Natl Acad Sci U S A* 97:8728–8733. [CrossRef Medline](#)
- Chennu N, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. *J Neurosci* 33:11194–11205. [CrossRef Medline](#)
- Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36:181–204. [CrossRef Medline](#)
- Clos M, Langner R, Meyer M, Oechslin MS, Zilles K, Eickhoff SB (2014) Effects of prior information on decoding degraded speech: an fMRI study. *Hum Brain Mapp* 35:61–74. [CrossRef Medline](#)
- Criaud M, Boulinguez P (2013) Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review. *Neurosci Biobehav Rev* 37:11–23. [CrossRef Medline](#)
- d'Acremont M, Schultz W, Bossaerts P (2013) The human brain encodes event frequencies while forming subjective beliefs. *J Neurosci* 33:10887–10897. [CrossRef Medline](#)
- de Lange FP, Rahnev DA, Donner TH, Lau H (2013) Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *J Neurosci* 33:1400–1410. [CrossRef Medline](#)
- De Martino B, Fleming SM, Garrett N, Dolan RJ (2013) Confidence in value-based choice. *Nat Neurosci* 16:105–110. [CrossRef Medline](#)
- Dienes Z (2008) Subjective measures of unconscious knowledge. *Prog Brain Res* 168:49–64. [CrossRef Medline](#)
- Downar J, Crawley AP, Mikulis DJ, Davis KD (2000) A multimodal cortical network for the detection of changes in the sensory environment. *Nat Neurosci* 3:277–283. [CrossRef Medline](#)
- Downar J, Crawley AP, Mikulis DJ, Davis KD (2002) A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *J Neurophysiol* 87:615–620. [Medline](#)
- Egner T, Monti JM, Summerfield C (2010) Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci* 30:16601–16608. [CrossRef Medline](#)
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335. [CrossRef Medline](#)
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716. [CrossRef Medline](#)
- Fetsch CR, Kiani R, Newsome WT, Shadlen MN (2014a) Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83:797–804. [CrossRef Medline](#)
- Fetsch CR, Kiani R, Shadlen MN (2014b) Predicting the accuracy of a decision: a neural mechanism of confidence. *Cold Spring Harb Symp Quant Biol* 79:185–197. [CrossRef Medline](#)
- Fiser J, Berkes P, Orbán G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn Sci* 14:119–130. [CrossRef Medline](#)
- Fleck MS, Daselaar SM, Dobbins IG, Cabeza R (2006) Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cereb Cortex* 16:1623–1630. [CrossRef Medline](#)
- Fleming SM, Dolan RJ (2012) Neural basis of metacognition. *Philos Trans R Soc B Biol Sci* 367:1338–1349. [CrossRef](#)
- Fleming SM, Lau HC (2014) How to measure metacognition. *Front Hum Neurosci* 8:1–9. [CrossRef Medline](#)
- Fleming SM, Maniscalco B, Ko Y, Amendi N, Ro T, Lau H (2015) Action-specific disruption of perceptual confidence. *Psychol Sci* 26:89–98. [CrossRef Medline](#)
- Friston K (2009) The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* 13:293–301. [CrossRef Medline](#)
- Galvin SJ, Podd JV, Drga V, Whitmore J (2003) Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon Bull Rev* 10:843–876. [CrossRef Medline](#)

- Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2009) Dynamic causal modeling of the response to frequency deviants. *J Neurophysiol* 101:2620–2631. [CrossRef Medline](#)
- Gherman S, Philiastides MG (2015) Neural representations of confidence emerge from the process of decision formation during perceptual choices. *Neuroimage* 106:134–143. [CrossRef Medline](#)
- Gilbert CD, Li W (2013) Top-down influences on visual processing. *Nat Rev Neurosci* 14:350–363. [CrossRef Medline](#)
- Grinband J, Hirsch J, Ferrera VP (2006) A neural representation of categorization uncertainty in the human brain. *Neuron* 49:757–763. [CrossRef Medline](#)
- Hampshire A, Chamberlain SR, Monti MM, Duncan J, Owen AM (2010) The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage* 50:1313–1319. [CrossRef Medline](#)
- Hebart MN, Schriever Y, Donner TH, Haynes JD (2016) The relationship between perceptual decision variables and confidence in the human brain. *Cereb Cortex* 26:118–130. [CrossRef Medline](#)
- Jiang J, Summerfield C, Egnér T (2013) Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *J Neurosci* 33:18438–18447. [CrossRef Medline](#)
- Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90:773–795. [CrossRef](#)
- Kepecs A, Mainen ZF (2012) A computational framework for the study of confidence in humans and animals. *Philos Trans R Soc Lond B Biol Sci* 367:1322–1337. [CrossRef Medline](#)
- Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455:227–231. [CrossRef Medline](#)
- Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324:759–764. [CrossRef Medline](#)
- Kiani R, Corthell L, Shadlen MN (2014) Choice certainty is informed by both evidence and decision time. *Neuron* 84:1329–1342. [CrossRef Medline](#)
- Kim H, Shimojo S, O'Doherty JP (2011) Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb Cortex* 21:769–776. [CrossRef Medline](#)
- Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP (2012a) Attention reverses the effect of prediction in silencing sensory signals. *Cereb Cortex* 22:2197–2206. [CrossRef Medline](#)
- Kok P, Jehee JF, de Lange FP (2012b) Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* 75:265–270. [CrossRef Medline](#)
- Kouider S, Long B, Le Stanc L, Charron S, Fievet AC, Barbosa LS, Gelskov SV (2015) Neural dynamics of prediction and surprise in infants. *Nat Commun* 6:8537. [CrossRef Medline](#)
- Larsson J, Smith AT (2012) fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb Cortex* 22:567–576. [CrossRef Medline](#)
- Law JR, Flanery MA, Wirth S, Yanike M, Smith AC, Frank LM, Suzuki WA, Brown EN, Stark CE (2005) Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. *J Neurosci* 25:5720–5729. [CrossRef Medline](#)
- Lebreton M, Abitbol R, Daunizeau J, Pessiglione M (2015) Automatic integration of confidence in the brain valuation signal. *Nat Neurosci* 18:1159–1167. [CrossRef Medline](#)
- Lee TS (2002) Top-down influence in early visual processing: a Bayesian perspective. *Physiol Behav* 77:645–650. [CrossRef Medline](#)
- Love J, Selker R, Marsman M, Jamil T, Drotmann D, Verhagen AJ, Ly A, Gronau QF, Smira M, Epskamp S, Matzke D, Wild A, Rouder JN, Morey RD, Wagenmakers EJ (2015) JASP (Version 0.7).
- Maloney LT, Dal Martello MF, Sahm C, Spillmann L (2005) Past trials influence perception of ambiguous motion quartets through pattern completion. *Proc Natl Acad Sci U S A* 102:3164–3169. [CrossRef Medline](#)
- Maniscalco B, Peters MA, Lau H (2016) Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Atten Percept Psychophys* 78:923–937. [CrossRef Medline](#)
- Mansouri FA, Tanaka K, Buckley MJ (2009) Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nat Rev Neurosci* 10:141–152. [CrossRef Medline](#)
- Meyniel F, Schlunegger D, Dehaene S (2015a) The sense of confidence during probabilistic learning: a normative account. *PLoS Comput Biol* 11:e1004305. [CrossRef Medline](#)
- Meyniel F, Sigman M, Mainen ZF (2015b) Confidence as Bayesian probability: from neural origins to behavior. *Neuron* 88:78–92. [CrossRef Medline](#)
- Morales J, Solovey G, Maniscalco B, Rahnev D, de Lange FP, Lau H (2015) Low attention impairs optimal incorporation of prior knowledge in perceptual decisions. *Atten Percept Psychophys* 77:2021–2036. [CrossRef Medline](#)
- Morris RW, Dezfouli A, Griffiths KR, Balleine BW (2014) Action-value comparisons in the dorsolateral prefrontal cortex control choice between goal-directed actions. *Nat Commun* 5:4390. [CrossRef Medline](#)
- Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2004) Dopamine neurons can represent context-dependent prediction error. *Neuron* 41:269–280. [CrossRef Medline](#)
- Nassar MR, Wilson RC, Healy B, Gold JI (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci* 30:12366–12378. [CrossRef Medline](#)
- Overgaard M, Sandberg K (2012) Kinds of access: different methods for report reveal different kinds of metacognitive access. *Philos Trans R Soc Lond B Biol Sci* 367:1287–1296. [CrossRef Medline](#)
- Pajani A, Kok P, Kouider S, de Lange FP (2015) Spontaneous activity patterns in primary visual cortex predispose to visual hallucinations. *J Neurosci* 35:12947–12953. [CrossRef Medline](#)
- Petrides M, Pandya DN (2002) Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur J Neurosci* 16:291–310. [CrossRef Medline](#)
- Petrusic WM, Baranski JV (2003) Judging confidence influences decision processing in comparative judgments. *Psychon Bull Rev* 10:177–183. [CrossRef Medline](#)
- Pleskac TJ, Busemeyer JR (2010) Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol Rev* 117:864–901. [CrossRef Medline](#)
- Rahnev D, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H (2011) Attention induces conservative subjective biases in visual perception. *Nat Neurosci* 14:1513–1515. [CrossRef Medline](#)
- Ratcliff R, Starns JJ (2009) Modelling confidence and response time in recognition memory. *Psychol Rev* 116:59–83. [CrossRef Medline](#)
- Reckless GE, Ousdal OT, Server A, Walter H, Andreassen OA, Jensen J (2014) The left inferior frontal gyrus is involved in adjusting response bias during a perceptual decision-making task. *Brain Behav* 4:398–407. [CrossRef Medline](#)
- Rorden C, Brett M (2000) Stereotaxic display of brain lesions. *Behav Neurol* 12:191–200. [CrossRef Medline](#)
- Sakagami M, Pan X (2007) Functional role of the ventrolateral prefrontal cortex in decision making. *Curr Opin Neurobiol* 17:228–233. [CrossRef Medline](#)
- Sandberg K, Timmermans B, Overgaard M, Cleeremans A (2010) Measuring consciousness: is one measure better than the other? *Conscious Cogn* 19:1069–1078. [CrossRef Medline](#)
- Schoenbaum G, Roesch M (2005) Orbitofrontal cortex, associative learning, and expectancies. *Neuron* 47:633–636. [CrossRef Medline](#)
- Seriès P, Seitz AR (2013) Learning what to expect (in visual perception). *Front Hum Neurosci* 7:1–14. [CrossRef](#)
- Seth AK, Dienes Z, Cleeremans A, Overgaard M, Pessoa L (2008) Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends Cogn Sci* 12:314–321. [CrossRef Medline](#)
- Sharot T, Korn CW, Dolan RJ (2011) How unrealistic optimism is maintained in the face of reality. *Nat Neurosci* 14:1475–1479. [CrossRef Medline](#)
- Sherman MT, Seth AK, Barrett AB, Kanai R (2015) Prior expectations facilitate metacognition for perceptual decision. *Conscious Cogn* 35:53–65. [CrossRef Medline](#)
- Sherman MT, Kanai R, Seth AK, VanRullen R (2016) Rhythmic influence of top-down perceptual priors in the phase of prestimulus occipital alpha oscillations. *J Cogn Neurosci* 28:1318–1330. [CrossRef Medline](#)
- Smith FW, Muckli L (2010) Nonstimulated early visual areas carry information about surrounding context. *Proc Natl Acad Sci U S A* 107:20099–20103. [CrossRef Medline](#)
- St. John-Saaltink E, Utzerath C, Kok P, Lau HC (2015) Expectation suppression in early visual cortex depends on task set. *10:e0131173*.
- Stalnaker TA, Cooch NK, Schoenbaum G (2015) What the orbitofrontal cortex does not do. *Nat Neurosci* 18:620–627. [CrossRef Medline](#)
- Summerfield C, de Lange FP (2014) Expectation in perceptual decision

- making: neural and computational mechanisms. *Nat Rev Neurosci* 15: 745–756. [CrossRef Medline](#)
- Suzuki M, Gottlieb J (2013) Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat Neurosci* 16:98–104. [CrossRef Medline](#)
- Tops M, Boksem MA (2011) A potential role of the inferior frontal gyrus and anterior insula in cognitive control, brain rhythms, and event-related potentials. *Front Psychol* 2:330. [CrossRef Medline](#)
- Trapp S, Bar M (2015) Prediction, context, and competition in visual recognition. *Ann N Y Acad Sci* 1339:190–198. [CrossRef Medline](#)
- Verbruggen F, Logan GD (2008) Response inhibition in the stop-signal paradigm. *Trends Cogn Sci* 12:418–424. [CrossRef Medline](#)
- Wacongne C, Labyt E, Van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci U S A* 108:1–6. [CrossRef Medline](#)
- Wallis JD (2007) Orbitofrontal cortex and its contribution to decision-making. *Annu Rev Neurosci* 30:31–56. [CrossRef Medline](#)
- White CN, Mumford JA, Poldrack RA (2012) Perceptual criteria in the human brain. *J Neurosci* 32:16716–16724. [CrossRef Medline](#)
- Wierchoń M, Paulewicz B, Asanowicz D, Timmermans B, Cleeremans A (2014) Different subjective awareness measures demonstrate the influence of visual identification on perceptual awareness ratings. *Conscious Cogn* 27C:109–120. [CrossRef Medline](#)
- Yeung N, Summerfield C (2012) Metacognition in human decision-making: confidence and error monitoring. *Philos Trans R Soc Lond B Biol Sci* 367:1310–1321. [CrossRef Medline](#)
- Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci* 10:301–308. [CrossRef Medline](#)
- Zylberberg A, Barttfeld P, Sigman M (2012) The construction of confidence in a perceptual decision. *Front Integr Neurosci* 6:79. [CrossRef Medline](#)